Multivariate and multiway analysis of hyperspectral and fluorescence landscape data

by

Helen Rutlidge

A thesis submitted for the

Degree of Doctor of Philosophy (Science)

University of Technology, Sydney

March, 2011

Dedicated to Michael Coleman and Peter Rutlidge

"For some life lasts a short while, but the memories it holds lasts forever."

Laura Swenson

Certificate of authorship and originality

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all the information sources and literature used are indicated in the thesis.

Helen Rutlidge

31/03/2011

Acknowledgements

Whilst this has been my project for the last five years there have a number of people who have contributed to the success of it. Firstly to Dr Brian Reedy, my supervisor, who first started me on this journey. Your guidance and encouragement have enhanced this process for me and made the project what it is.

I would like to thank Professor Peter Ralph for his suggestion and assistance with the algal cultures in this project. Thank you to Vinod Kumar and Penelope Ajani for their technical assistance with the algal cultures.

I thank Dr Dominic Hare and Christine Austin for all of their assistance with laser ablation inductively coupled plasma. Both of you were always willing to answer any questions I had and provide assistance when required.

I would also like to Fiona Burger, Jonathan Mak and Anne Rutlidge for collecting soil samples for me.

Thanks also go to everyone in Office 4.60 and other fellow students (past and present) who I have shared this journey with. You have always been willing to celebrate the good times and find the silver lining in the bad. All of you have made being at uni enjoyable.

I would like to thank my family and friends for their support and understanding whilst I have been at university. Thank you also for making sure I had a life outside my PhD.

Table of contents

CHAPTER 1: INTRODUCTION **2**

1.1 HYPERSPECTRAL IMAGING 2

1.1.1 UV-visible imaging *4*

1.1.1.1 Fluorescence imaging 6

1.1.2 Infrared imaging techniques *6*

1.1.2.1 Near infrared imaging 7

1.1.2.2 Mid-infrared mapping and imaging 9

1.1.3 Raman mapping and imaging *12*

1.1.4 Molecular mass spectrometry imaging *14*

1.1.5 Elemental mapping *16*

1.1.5.1 X-ray mapping techniques 16

1.1.5.2 Laser ablation inductively coupled mass spectroscopy 17

1.2 ANALYSIS OF HYPERSPECTRAL IMAGES 18

1.2.1 Preprocessing methods *19*

1.2.1.1 Savitzky-Golay derivatives 20

1.2.1.2 Mean-centring 20

1.2.1.3 Variance scaling 20

1.2.1.4 Autoscaling 20

1.2.1.5 Vector normalisation 21

1.2.2 Clustering techniques *21*

1.2.2.1 Distance measures 21

1.2.2.2 Hierarchical cluster analysis 22

1.2.2.3 Fuzzy C-means clustering 24

1.2.2.4 Self-organising feature map 25

1.2.3 "Factor analysis" techniques *26*

1.2.3.1 Principal component analysis 27

1.2.3.2 Partial least squares analysis 28

1.2.4 End member techniques 30

1.2.4.1 Spectral angle mapper 30

1.2.4.2 Spectral feature fitting 31

1.2.4.3 Mixture tuned matched filter 32

1.2.5 Multiway techniques 33

1.2.5.1 PARAFAC 33

1.2.5.2 Tucker models 34

1.2.5.3 Application of multiway techniques on hyperspectral images 34

1.2.6 Dimension reduction strategies 35

1.3 PROJECT AIMS 35

CHAPTER 2: CLASSIFICATION OF HETEROGENEOUS SOLIDS USING INFRARED HYPERSPECTRAL IMAGING 37

2.1 INTRODUCTION 37

2.1.1 Hyperspectral imaging 37

2.1.2 Dimension reduction strategies 38

2.1.2.1 PCA on PC1 38

2.1.2.2 ASMULIM 39

2.1.2.3 MOLMAP 39

2.1.3 Classification techniques 39

2.1.3.1 Discriminant analysis 40

2.1.3.2 Soft independent modelling of class analogy 40

2.1.4 Aims 41

2.2 EXPERIMENTAL AND DATA ANALYSIS METHODS 42

2.2.1 Sample preparation 42

2.2.2 Instrumental 43

2.2.3 Data preprocessing 44

2.2.4 Data analysis *44*

2.2.4.1 ASD and MIQ methods 44

2.2.4.2 Average and median methods 46

2.2.4.3 Clustering methods 46

2.2.4.4 PCA on PC1 47

2.2.4.5 ASMULIM 48

2.2.4.6 Endmember techniques 48

2.3 RESULTS AND DISCUSSION 49

2.3.1 Choice of compounds *49*

2.3.2 Sampling considerations *49*

2.3.3 Particle size vs pixel size *50*

2.3.4 Comparison of classification approaches *51*

2.3.5 Endmember classification results *56*

2.4 CONCLUSIONS 57

CHAPTER 3: CLASSIFICATION OF SUBURBAN SOIL SAMPLES USING HYPERSPECTRAL IMAGING ... 59

3.1 INTRODUCTION 59

3.1.1 Soil *59*

3.1.1.1 Forensic analysis of soil 59

3.1.1.2 Mid-infrared analysis of soil 60

3.1.1.3 Elemental analysis of soil 60

3.1.2 LA-ICP-MS *62*

3.1.2.1 LA-ICP-MS calibration 62

3.1.3 Overall aims *63*

3.2 EXPERIMENTAL AND DATA ANALYSIS METHODS 64

3.2.1 Preliminary study *64*

3.2.1.1 Sample preparation 64

3.2.1.2 Infrared imaging 65

3.2.1.3 LA-ICP-MS elemental imaging......65

3.2.1.4 Data analysis......67

3.2.2 Final study......67

3.2.2.1 Sample preparation......67

3.2.2.2 Element selection......68

3.2.2.3 LA-ICP-MS elemental imaging......69

3.2.2.4 Data analysis......70

3.3 RESULTS AND DISCUSSION......71

3.3.1 LA-ICP-MS normalisation......71

3.3.2 Preliminary soil sample set......72

3.3.3 Final soil sample set......76

3.3.3.1 Element selection......76

3.3.3.2 Classification results......78

3.3.3.3 Comparison of different imaging approaches......81

3.4 CONCLUSIONS......82

CHAPTER 4: PIGMENT ANALYSIS IN ALGAE USING FLUORESCENT LANDSCAPES AND PARAFAC...85

4.1 INTRODUCTION......85

4.1.1 Algae......85

4.1.1.1 Pigment analysis......85

4.1.1.2 Algal fluorescence......86

4.1.1.3 Adaptation of algae to light......87

4.1.2 Fluorescent landscapes......88

4.1.3 PARAFAC......89

4.1.4 Aims......89

4.2 EXPERIMENTAL AND DATA ANALYSIS METHODS......90

4.2.1 Samples......90

4.2.2 Light table......90

4.2.3 Preliminary study *91*

4.2.4 Final study *91*

4.2.5 Measurements *92*

4.2.5.1 Fluorescence landscapes 92

4.2.6 Linearity Studies *93*

4.2.7 Data analysis *93*

4.3 RESULTS AND DISCUSSION 94

4.3.1 Preliminary investigations *94*

4.3.1.1 Algal landscapes 94

4.3.1.2 Linearity 95

4.3.2 Preliminary study *97*

4.3.2.1 Chlorophyll a concentration 97

4.3.2.2 PARAFAC results 98

4.3.3 Final study *102*

4.3.3.1 Chlorophyll a concentration 102

4.3.3.2 Yield measurements 103

4.3.3.3 PARAFAC results 103

4.4 CONCLUSIONS 107

CHAPTER 5: APPLICATION OF PARAFAC TO LATENT FINGERPRINT VISUALISATION 109

5.1 INTRODUCTION 109

5.1.1 Fingerprints on difficult surfaces *110*

5.1.1.1 Hyperspectral imaging 110

5.1.1.2 Fluorescence imaging 112

5.1.1.3 Nanoparticles 114

5.1.1.4 Other techniques 115

5.1.2 Background correction *116*

5.1.3 Aims *116*

5.2 EXPERIMENTAL AND DATA ANALYSIS METHODS 117

5.2.1 Samples 117

5.2.1.1 Drink can 117

5.2.1.2 Polymer banknotes 117

5.2.1.3 Glossy printed card 118

5.2.2 Instrumental 118

5.2.2.1 Polilight / Poliview system 118

5.2.2.2 HSI Examiner™ 100 QD 119

5.2.3 Data preprocessing 120

5.2.3.1 Polilight / Poliview system 120

5.2.3.2 HSI Examiner™ 100 QD 120

5.2.4 Data analysis 120

5.2.4.1 PARAFAC 120

5.2.4.2 PCA 121

5.2.4.3 MIA approach 121

5.3 RESULTS AND DISCUSSION 121

5.3.1 Aluminium drink can 121

5.3.1.1 PARAFAC 123

5.3.1.2 PCA 126

5.3.1.3 MIA approach 127

5.3.2 Banknote (Polilight) 128

5.3.2.1 PARAFAC 129

5.3.2.2 PCA 130

5.3.2.3 MIA approach 132

5.3.3 Banknote (HSI Examiner™) 133

5.3.3.1 PARAFAC 135

5.3.3.2 PCA 136

5.3.4 Glossy printed card (Polilight) 137

5.3.4.1 PARAFAC.......137

5.3.4.2 PCA.......140

5.3.5 Glossy Card (HSI Examiner™).......140

5.3.5.1 PARAFAC.......141

5.3.5.2 PCA.......141

5.3.5.3 MIA approach.......142

5.3.6 General Discussion.......143

5.3.6.1 Comparison of PARAFAC and PCA.......143

5.3.6.2 Background interference.......143

5.3.6.3 MIA approach.......146

5.4 CONCLUSIONS.......147

CHAPTER 6: CONCLUSIONS AND FURTHER WORK.......149

List of figures

Figure 1-1 Diagrammatical representation of an hyperspectral data cube (image from Tahtouh et al.[2]). 2

Figure 1-2 A portion of the electromagnetic spectrum, showing representative molecular processes that occur when light in each region is absorbed (image adapted from Harris[4]). 3

Figure 1-3 Approximate stretching vibrations in the MIR region.[8] 9

Figure 1-4 Classification of methods for hyperspectral image analysis (adapted from Gendrin et al.[66]). 19

Figure 1-5 An example of a dendrogram from HCA 24

Figure 1-6 Schematic of the SOFM training process.[245] 26

Figure 1-7 The MIA approach (A) The scatter score plot (PC1 vs PC2). (B) The defined classes in the score plot of (A). (C) The classes in the original image.[270] 28

Figure 1-8 Diagrammatical representation of PLS, where X is the spectral data, c is the reference values, T is the common scores matrix, P and q are loading matrices, and E and f are error matrices.[247] 29

Figure 1-9 An illustration of the difference in the way that SAM (left panel) and SFF (right panel) compare a known endmember spectrum to an unknown pixel spectrum.[277] 32

Figure 1-10 Diagrammatical representation of PARAFAC. 33

Figure 2-1 Schematic of the general approach to dimension reduction and classification of the resulting feature vectors. 42

Figure 2-2 Schematic of data treatment for the MIQ method (data is treated analogously in the ASD method); "M1 S1" is Mixture 1-Sample1, etc. 45

Figure 2-3 Construction of median-interquartile range "super-spectrum" used in the MIQ method. 46

Figure 2-4 Conversion of a cluster image (such as from HCA) to a feature vector for an image; this vector is simply a histogram of the number of pixels in the image that belong to each cluster....47

Figure 2-5 Representative spectra from a reflectance FTIR spectral image of Mixture 1 (25% w/w of each anthracene, anthraquinone, aspirin, and citric acid)....50

Figure 2-6 Two perspectives of the discriminant scores plot obtained by classifying scores from the PCA of the single-point spectra....53

Figure 2-7 Two perspectives of the discriminant scores plot obtained by classifying scores from the PCA of the MIQ super-spectra from the images....54

Figure 3-1 PCA scores plot for LA-ICP-MS data using silicon as IS....71

Figure 3-2 PCA scores plot for LA-ICP-MS elemental ratios....72

Figure 3-3 Average MIR spectra for the soil locations....74

Figure 3-4 LDA scores plot for MIQ of MIR image data (filled circles represent the training samples and open ciricles test samples)....75

Figure 3-5 LDA scores plot for MIQ of LA-ICP-MS elemental maps (filled circles represent the training samples and open ciricles test samples)....75

Figure 3-6 LDA scores plot for MIQ of LA-ICP-MS elemental maps from final sample set (filled circles represent the training samples and open ciricles test samples)....80

Figure 3-7 PCA scores plot for MIQ of LA-ICP-MS elemental maps from final sample set (filled circles represent the training samples and open ciricles test samples)....81

Figure 4-1 A example of a fluorescent landscape....88

Figure 4-2 Fluorescent landscape for Chaetoceros sp.....95

Figure 4-3 Fluorescent landscape for Dunaliella tertiolecta....95

Figure 4-4 Fluorescence intensity vs. chlorophyll a concentration at chlorophyll a peak for Chaetoceros sp.....96

Figure 4-5 Fluorescence intensity vs. chlorophyll a concentration at chlorophyll a peak for Dunaliella tertiolecta....97

Figure 4-6 Chlorophyll a concentration measured using Trilogy Fluorometer in the preliminary study....98

Figure 4-7 PARAFAC components in preliminary study. 99

Figure 4-8 PARAFAC scores for component 1 in preliminary study....100

Figure 4-9 PARAFAC scores for component 2 in preliminary study....100

Figure 4-10 PARAFAC scores for component 3 in preliminary study....101

Figure 4-11 PARAFAC scores for component 4 in preliminary study....101

Figure 4-12 Chlorophyll a concentration measured using Trilogy Fluorometer in the final study....102

Figure 4-13 PARAFAC components in final study. 103

Figure 4-14 PARAFAC scores for component 1 in final study....105

Figure 4-15 PARAFAC scores for component 2 in final study....105

Figure 4-16 PARAFAC scores for component 3 in final study....106

Figure 4-17 PARAFAC scores for component 4 in final study....106

Figure 5-1 Fluorescent image of ethyl cyanoacrylate-fumed fingerprint on aluminium drink can stained with Rhodamine 6G (λ_{ex} 490 nm, λ_{em} 555 nm)....122

Figure 5-2 Average fluorescent landscapes from different regions of aluminium drink can..122

Figure 5-3 PARAFAC score images from ethyl cyanoacrylate-fumed fingerprint on aluminium drink stained with Rhodamine 6G. 124

Figure 5-4 PARAFAC score images after mean centring of fluorescent landscapes from ethyl cyanoacrylate-fumed fingerprint on aluminium drink stained with Rhodamine 6G....125

Figure 5-5 PCA score images from ethyl cyanoacrylate-fumed fingerprint on aluminium drink stained with Rhodamine 6G. 126

Figure 5-6 PARAFAC scores scatter plot with the selected pixels green. 127

Figure 5-7 MIA image from PARAFAC of aluminium drink can.. 128

Figure 5-8 MIA image from PCA of aluminium drink can. 128

Figure 5-9 Fluorescent image from Polilight of fresh fingerprint on banknote powered with Brilliant Yellow (λ_{ex} = 450 nm, λ_{em} = 530 nm). 129

Figure 5-10 PARAFAC score images after mean centring of fluorescent landscapes on banknote powdered with Brilliant Yellow (Polilight). 131

Figure 5-11 PCA score images after mean centring of fluorescent landscapes from fingerprint on banknote powdered with Brilliant Yellow (Polilight). 131

Figure 5-12 MIA image from PARAFAC of banknote.. 132

Figure 5-13 MIA image from PARAFAC of banknote showing ridge detail over black intaglio printing. 132

Figure 5-14 Fluorescent image from HSI Examiner™ of fresh fingerprint on banknote powered with Brilliant Yellow (λ_{ex} = 315 – 400 nm, λ_{em} = 510 nm). 133

Figure 5-15 PARAFAC score images after mean centring of fluorescent landscapes from fingerprint on banknote powdered with Brilliant Yellow (HSI Examiner™). 134

Figure 5-16 Emission loadings from PARAFAC of banknote (HSI Examiner™)........................ 135

Figure 5-17 PCA score images after mean centring of fluorescent landscapes from fingerprint on banknote powdered with Brilliant Yellow (HSI Examiner™)... 136

Figure 5-18 Fluorescent image from Polilight of ethyl cyanoacrylate-fumed fingerprint on glossy printed card powered with Brilliant Orange (λ_{ex} = 490 nm & λ_{em} = 610 nm). 137

Figure 5-19 PARAFAC score images after mean centring of fluorescent landscapes from fingerprint on glossy card powdered with Brilliant Orange (Polilight). 138

Figure 5-20 PCA score images after mean centring of fluorescent landscapes from fingerprint on glossy card powdered with Brilliant Orange (Polilight)....139

Figure 5-21 Fluorescent image from HSI Examiner™ of ethyl cyanoacrylate fumed fingerprint on glossy printed card powered with Brilliant Orange (λ_{ex} = 450-500 nm, λ_{em} = 590 nm)......140

Figure 5-22 PARAFAC score images after mean centring of fluorescent landscapes from fingerprint on glossy card powdered with Brilliant Orange (HSI Examiner™)....141

Figure 5-23 PCA score images after mean centring of fluorescent landscapes from fingerprint on glossy card powdered with Brilliant Orange (HSI Examiner™)....142

Figure 5-24 MIA image from PARAFAC of glossy card (HSI Examiner™)....142

Figure 5-25 Comparison of fluorescent powder emission spectra and PARAFAC loadings for HSI Examiner™ datasets....144

Figure 5-26 Comparison of fluorescent powder emission spectra and PARAFAC loadings for Polilight datasets....145

Figure 5-27 Average emission spectra from 400-450 nm excitation of banknote using HSI Examiner™....146

List of tables

Table 2-1 Compositions of the four powder mixtures. 43

Table 2-2 Summary of results for image classification methods. 52

Table 3-1 Suburban soils collected for preliminary study. 64

Table 3-2 Experimental parameters for LA-ICP-MS analysis. 66

Table 3-3 Groups for calculation of elemental ratios in preliminary study. 66

Table 3-4 Suburban soils collected for final study. 68

Table 3-5 Experimental parameters for LA-ICP-MS element selection. 69

Table 3-6 Groups for calculation of elemental ratios in final study. 70

Table 3-7 LDA correct classification percentages for preliminary soil sample set. 73

Table 3-8 ANOVA and Kruskal-Wallis test results for element selection study (N.B. sig. < 0.002 used due to the number of variables). 77

Table 3-9 LDA correct classification percentages for final soil sample set. 79

Table 4-1 Kodak Wratten filters used in light table for algal studies. 91

Table 4-2 Experimental parameters for Cary Eclipse Spectrophotometer in preliminary study. 92

Table 4-3 Experimental parameters for Cary Eclipse Spectrophotometer in final study. 93

Table 4-4 Dilution series for linearity studies. 93

Table 5-1 Main constituents of latent fingerprint deposits.[429] 109

Table 5-2 Wavelengths collected for each sample using the Polilight / Poliview system. 119

Table 5-3 Experimental parameters for collection of fluorescent images using HSI Examiner™. 119

Abbreviations

ASD	average-standard deviation
ASMULIM	analysis of a set of multivariate images
ATR	attenuated total reflection
chl	chlorophyll
DA	discriminant analysis
DFO	1,8-diazafluorene-9-one
FPA	focal plane array
HCA	hierarchical cluster analysis
HPLC	high pressure liquid chromatography
IS	internal standard
LA-ICP-MS	laser ablation inductively coupled plasma mass spectrometry
LOO	leave-one-out
MIA	multivariate image analysis
MIQ	median-interquartile range
MIR	mid-infrared
NIR	near-infrared
PARAFAC	parallel factor analysis

PCA	principal component analysis
SAM	spectral angle mapper
SIMCA	soft independent modelling of class analogy
SOFM	self organising feature map
UV	ultraviolet

Abstract

The majority of the methods that have been used for the analysis of hyperspectral images have focussed on classification of spectra within an image, and there have been few examples of using whole hyperspectral images (one image per sample) for classification purposes. In this project, feature vectors were created to capture the heterogeneity present in each hyperspectral image for subsequent classification. All of the feature vectors methods were compared to single-point spectra and literature methods. The MIQ (the median-interquartile superspectrum) achieved the best classification of a set of heterogeneous powder mixtures, classifying 100% of the external test set.

Mid-infrared (MIR) hyperspectral images and LA-ICP-MS elemental maps were collected of 'real-world' samples of soil collected from various locations around Sydney. For both the MIR and LA-ICP-MS data, the MIQ feature vector achieved superior classification compared to the averages alone (used to represent single-point spectra). For the Sydney soil samples, the classification performance of the LA-ICP-MS elemental maps was superior to MIR reflection hyperspectral images. For the LA-ICP-MS data, using elemental ratios for calibration proved to be a more suitable strategy than using silicon as an internal standard.

A major application of PARAFAC has been the analysis of fluorescent landscapes. In the other main area of this project, novel applications of PARAFAC were investigated. The first application explored was the use of PARAFAC and fluorescent landscapes for the characterisation of phytoplankton. Algal species were placed under different wavelengths of light and monitored over time to determine their response to these environments. PARAFAC was able to isolate components that were identified as β-carotene, chlorophyll *a* and *b*, and chlorophyll *a* degradation products. The advantage of using PARAFAC for the characterisation of phytoplankton is that it can simultaneously determine all pigment concentrations without the need for prior separation or extraction.

Background interference due to highly coloured / patterned and fluorescence remains an issue with many techniques of fingerprint visualisation. The second application of PARAFAC in this project was to analyse fluorescent landscapes constructed by collecting series of images of fluorescent fingerprints on these traditionally difficult backgrounds. Whilst PARAFAC successfully isolated the fluorescence due to the fingerprint, there was typically only a small reduction in background interference. However, it is expected that PARAFAC would be more successful for samples where the background is also fluorescent, such as DFO-developed prints on fluorescent yellow paper.